

ML Academy - Project 1 Overview

Abstract:

The main goal of this project is to make you feel familiar with Supervised Learning Algorithms like Regression and Gradient Boosting algorithms along with practicing more Exploratory Data Analysis techniques and developing your ML intuition.

Description:

The Tunisian Company of Electricity and Gas (STEG) is a public and a non-administrative company, it is responsible for delivering electricity and gas across Tunisia. The company suffered tremendous losses in the order of 200 million Tunisian Dinars due to fraudulent manipulations of meters by consumers.

Using the client's billing history, the aim of the challenge is to detect and recognize clients involved in fraudulent activities.

The solution will enhance the company's revenues and reduce the losses caused by such fraudulent activities.

Data Problem:

Building a model that will help classify which customer is likely to commit fraud and by that saving the company from making losses.

Project Steps:

The project will be hold in 3 phases:

- **Exploratory Data Analysis (EDA)** phase, where you are supposed to work on understanding, clearing, sorting and extracting important features from the data provided.
Estimated Task Time: 2 days
- **Modeling** phase, where you will be able to use all the work done in EDA in order to create your own ML model (or models) using regression techniques.

Estimated Task Time: 2days

Submission:

Make sure that you fill this form on the deadline: <https://forms.gle/ay5ZBZfLUT6A2t8J7>, and that you gave access permissions to all the trainers on your final Colab.

Materials:

- **Data:**

(<https://drive.google.com/drive/folders/1lXq9qHd-xGFxEzMvOrPe6AkExuUn8XMZ?usp=sharing>)

This drive folder contains the dataset splitted into train and test sets along with a Readme txt file describing all the necessary information you need to know about the variables.

- **Useful Resources:**

- <https://towardsdatascience.com/exploratory-data-analysis-8fc1cb20fd15>
- <https://towardsdatascience.com/exploratory-data-analysis-in-python-c9a77dfa39ce>
- <https://machinelearningmastery.com/gentle-introduction-gradient-boosting-algorithm-machine-learning/>
- <https://towardsdatascience.com/understanding-gradient-boosting-machines-9be756fe76ab>

Rules:

- Do not cheat! You are here to learn
- Your work will be evaluated as a team
- We will evaluate and check your progress on each phase of the assignment
- Final evaluation be mainly on:
 - 20% Quality of data after EDA Phase
 - 25% Modeling
 - 25% Accuracy of predictions
 - 15% Quality of Code
 - 15% Team Work